DOCUMENT RESUME

ED 041 038 TM 000 002

AUTHOR Garvin, Alfred D.
TITLE The Applicability of Criterion-Referenced Measurement by Content Area and Level.
PUB DATE Mar 70
NOTE 8p.; From symposium "Criterion-Referenced Measurement: Emerging Issues" (Joint annual meetings of the American Educational Research Association and National Council on Measurement in Education, Minneapolis, Minn., March, 1970)

EDRS PRICE EDRS Price MF-$0.25 HC-$0.50
DESCRIPTORS *Criterion Referenced Tests, Evaluation, *Evaluation Methods, Measurement, *Measurement Techniques, *Norm Referenced Tests, Symposia, *Testing, Test Selection

ABSTRACT

In choosing between criterion-referenced and norm-referenced measurement strategy we should consider the nature of the decisions to be based on the resulting scores. If the decision involves selecting some fixed quota from the high (or low) end of an available competence continuum, then norm-referenced measurement is indicated. If, however, the decision involves certifying the attainment of some "a priori" standard of competence whether in some practitioner field or in some tool-skill academic field, then criterion-referenced measurement is indicated. In short, the choice between these two strategies should reflect the relative importance of quotas and standards in these decisions. It is suggested that the relative applicability of these strategies varies across content areas from the Humanities (norm-referenced) to the applied physical science professions (criterion-referenced). (Author)

ED041038

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

The Applicability of Criterion-Referenced Measurement* by Content Area and Level

Alfred D. Garvin
University of Cincinnati

No matter how constructive the topic that is specified before the colon in its title, any symposium on "Emerging Issues" is built around deliberately selected differences of opinion. A symposium on the topic "Peace on Earth: Emerging Issues" might very well end in a fist fight. We can expect that the individual papers in any such symposium will be models of internal consistency; the crucial issues will emerge between successive papers rather than within them. Thus, if Dr. Ebel will permit me, I must comment briefly on his paper in order to define the issue that emerges as I follow it with mine.

Dr. Ebel's paper dealt with whether we should use Criterion-Referenced Measurement (CRM) or Norm Referenced Measurement (NRM); mine deals with when each should be used. Although he clearly favored one above the other, his approach suggested that, in any given case, there was a choice that could be made. I acknowledge that this grossly oversimplifies his views. I will not oversimplify my own by stating merely that there never is a choice at all. The position that I take is this: In certain cases, CRM is irrelevant because, in fact, no meaningful criterion applies. In these cases, NRM must be used if there is to be any measurement at all. However, there are other cases where a meaningful criterion is inherent in the instructional objectives of the unit involved. If one measures the outcomes of such a unit at

*In Robert Glaser (Chm.) "Criterion-Referenced Measurement: Emerging Issues." Symposium presented at a joint senssion of the AERA-NCME annual meetings, Minneapolis, Minnesota, March, 1970.

TM000002

all, he is, in fact, conducting CRM. Between these two extremes, we might posit a continuum of relevance between criteria and instructional objectives.

The thesis I advance here may be summarized as follows: Our primary concern is with measuring the attainment of instructional objectives. The relevance of meaningful criteria to these instructional objectives dictates both the possibility of, and the necessity for, CRM. The relevance of criteria to instructional objectives is inherent in the content (and the level) of the instructional unit involved. Thus, for any given unit of instruction, we are not free to choose between CRM and NRM.

The basic issue that emerges here is, of course, the tenability of this thesis. In order to defend it, I will need a running start. I want to say some very fundamental things about instructional objectives, criteria, and measurement, per se, before presuming to prescribe the measurement technique to use for any given unit of instruction.

First of all, measurement is not an end in itself; we do not conduct instruction just to measure its effect. Furthermore, the process of instruction is not an end in itself; the process is intended to accomplish something. A classical behaviorist might say that the general objective of all instruction is to change the a priori probabilities among response alternatives in an anticipatible situation. A hard-nosed pragmatist might say that the objective of instruction is to get something done--and done right! I offer this account: The objective of instruction is to cause a change in some modifiable trait within the individual instructed. The trait involved

may be his knowledge of certain facts of U.S. History, his understanding of Boyle's law in physics, or his ability to translate selected passages from German into English. It may be his attitude toward all ethnic minorities, his belief in reincarnation, or his taste in literature. It may be his skill in parking trucks, playing tennis, or pulling teeth. If we add a psychomotor domain to the better-known cognitive and affective domains, we substantially anticipate most of the traits commonly specified in our instructional objectives.

Some of these traits are modifiable in degree; it is meaningful to speak of one having more or less of it and, by the conventions of trait-naming, having more of it is generally considered to be better than having less of it. Other traits are modifiable only in kind; changes in these traits are qualitative rather than quantitative. Most of these are comprehended in the affective domain of instructional objectives. While these may be as important as the quantitative traits, I will defer consideration of them here. A statement of instructional objectives must specify the desired final state of the trait or traits involved within the individual instructed. This may be the maximum level of which he is capable or it may be desired that he attain some predetermined level of this trait.

The necessity for any form of measurement at all arises in the fact that, ultimately, someone is going to do something about the extent to which different individuals attained these instructional objectives. This someone, or another someone, may also want to do something about the instructional process itself and/or those who conduct it. The primary purpose of

measurement is to inform the decisions these somebodies must make.

There are two ways to measure the final state of any trait of interest in a group--and these two ways apply to any quantitative variable. We can compare the trait-levels attained by two or more individuals with each other or we can compare each such level with some "standard level." The first of these procedures is an operational definition of NRM; the second, of CRM. As a practical matter, it must be recognized that these traits are merely psychological constructs; to the extent that they have any being at all, they exist in the neural organization of the individual. The point is that they cannot be measured directly. What we do, of course, is this: We contrive a set of tasks (i.e., test items) that we judge to be valid behavioral correlates of the trait of interest. Next, we either rate performances on this set of tasks or simply count "successes" on some arbitrary basis. Then we take the score resulting from this process as a "measure" of this underlying trait. As we all know full well, this is easier to do for some traits than for others. Nevertheless, whether we are using NRM or CRM, we contrive a set of tasks that embody, in behavioral terms, the instructional objectives of the unit.

To the extent that these contrived, classroom tasks correspond to some subsequent, extra-classroom task that must be performed at some "standard," i.e., criterion, level of proficiency in at least somes situations, CRM is possible. Of course, this does not mean that it is desirable nor that it is feasible. We may turn now to a consideration of those extra-classroom tasks that might provide a "meaningful Criterion" for our classroom tests.

There are certain tasks that, by their very nature, must be performed at a specifiably high level in almost every imaginable situation. Landing an airliner at O'Hare field is one; compounding a prescription is another. Any task in which public safety is involved falls in this category. There are other tasks in which some latitude of competence is permissible, even though a "criterion" level could be specified. No one would be seriously hurt if these were done less than perfectly and, in general, a deficient performance could be remedied. Cooking, housepainting, translating Latin, balancing checkbooks, and spelling fall in this category. There are tasks in which several different levels of performance are acceptable in as many different situations. There is a market for several different typing speeds and one might translate foreign documents in minutes or in days. There are some tasks that need not be done to <u>any</u> standard. There is room in this world for third-rate poets, inept actors, and simply awful golfers. All of these abilities are acquired through instruction. To the extent that a "predetermined," i.e., criterion, level of performance in these tasks is crucial, the tests on such instruction ought to be criterion-referenced.

There is a class of instructional objectives in which the extra-classroom task envisioned is to be performed in the next classroom. Many units of instruction are intended primarily to prepare the individual to undertake the next unit in the sequence. To the extent that it is meaśonable to specify an entering level of competence for this next unit, this level is a meaningful criterion for the present unit, whether or not the next unit is, itself, criterion-oriented. This is true in any cumulative content area. Mathematics and foreign languages are excellent examples.

There are two more things that must be said about "standards" or criteria--they arise from tasks performed outside the classroom, but they are not independent of the capabilities displayed within the classroom. An arbitrary standard of performance specified by the instructor is not a criterion, as I use the term. For purposes of his own, he may require that his students diagram four out of five selected sentences correctly or recite all the capitals of Europe in alphabetical order in one minute. These are not meaningful criteria. Requiring a correct diagnosis from a standard set of symptoms is. Next, a meaningful criterion must lie within the range of capabilities of those available to perform the task involved. It is pointless to demand prodigious reading speed for entry into third grade or to rate all piano students against a Horowitz recording. As a practical matter, criteria evolve from performance data gathered by NRM. The "predetermined" levels of performance the "real-world" requires in its important tasks are predetermined by available competence.

Before suggesting some general rules for matching measurement techniques with content areas and levels thereof, it would be well to reflect on the ultimate purpose of measurement--to inform decision making. Decisions must be made about individuals and decisions must be made about tasks. If we must select a fixed quota from, say, the top of some available distribution of relevant ability, no matter how high or low this "top" level may be, NRM is indicated. If we must select individuals to perform a given task at some fixed standard of competence, no matter how many or how few qualify, then CRM is indicated. As previously explained, standards tend to accomodate

available quotas, and the important work of the world does get done with the kinds of people there are.

When we apply the rationale developed above to the entire range of activities subsumed in the term "instruction," some general principles emerge regarding the applicability of CRM to various content areas and the various levels of these.

1. Unless at least one of the instructional objectives of a unit envisions a task that must subsequently be performed at a specified level of competence in at least some situation, CRM is irrelevant because there is no criterion. In this sense, the entire sequence of "social studies" provides no meaningful criterion except, possibly, the entry level for certain "honors" courses.

2. If public safety, econmic responsibility, or other ethical considerations demand that certain tasks be performed only be those "qualified" for them by formal instruction, then CRM of the outcomes of such instruction is clearly indicated. The criterion here is the licensing standards of the profession involved. All professional instruction in the medical arts, law, finance, engineering, and the applied physical and social sciences generally is clearly in this category. Teaching--at any level--ought to be. However, entry to such professional training is typically based on NRM since training capacity imposes a "quota."

3. In any instructional sequence where the content is inherently cumulative and the rigor progressively greater, CRM should be used to control entry to successive units. However, if there are several different sequences, differing widely in rigor, NRM is more useful in making appropriate placements. The best examples of these are mathematics and the physical and biological sciences in secondary school. Reading is the definitive example in the elementary grades.

4. There are certain content areas to which criteria do apply but not everyone need meet them. These are the "required subjects; everyone must try to learn them--if only as a matter of public policy--but it is almost preordained that some of them will not. Home economics and physical education are relatively non-controversial examples at the secondary level; at the college level, these become professions and CRM applies.

At the outset of this paper, I said it would raise issues. I may live to regret it, but I must raise just one more. According to my rationale, English is a subject that not everyone need master. If my thesis can survive this outrageous implication, it can survive anything.